Published in final edited form as:

Annu Rev Biophys. 2017 May 22; 46: 105–130. doi:10.1146/annurev-biophys-070816-033639.

Rate Constants and Mechanisms of Protein–Ligand Binding

Xiaodong Pang^{1,2} and Huan-Xiang Zhou^{1,2}

¹Department of Physics, Florida State University, Tallahassee, Florida 32306

²Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306

Abstract

Whereas protein–ligand binding affinities have long-established prominence, binding rate constants and binding mechanisms have gained increasing attention in recent years. Both new computational methods and new experimental techniques have been developed to characterize the latter properties. It is now realized that binding mechanisms, like binding rate constants, can and should be quantitatively determined. In this review, we summarize studies and synthesize ideas on several topics in the hope of providing a coherent picture of and physical insight into binding kinetics. The topics include microscopic formulation of the kinetic problem and its reduction to simple rate equations; computation of binding rate constants; quantitative determination of binding mechanisms; and elucidation of physical factors that control binding rate constants and mechanisms.

Keywords

binding rate constant; binding mechanism; conformational selection; induced fit; dock and coalesce; reactive flux

INTRODUCTION

The stereospecific binding of proteins with small molecules and macromolecular targets, including other proteins, nucleic acids, and their complexes, is at the core of the many biochemical processes essential to life. The magnitudes of the binding rate constants, not necessarily the magnitudes of the binding affinities, may dictate the fates of the biochemical processes because the latter are often under kinetic rather than thermodynamic control in their cellular contexts. Unimolecular reactions are usually modeled by a single reaction coordinate representing key intramolecular motions, or by a reaction coordinate together with one or more orthogonal coordinates when they are tightly coupled (3, 101). In contrast, binding reactions are bimolecular and involve both intermolecular relative motions and intramolecular motions. Accordingly, the task of theorists to reduce binding reactions to a binding rate constant in a microscopic description is challenging. Meanwhile, the interplay between intermolecular and intramolecular motions during the binding process can give rise

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

to a great variety of binding mechanisms, as represented by major pathways via which the reactant molecules proceed to yield the native complex. Binding pathways not only provide physical insight into binding reactions but may also have practical value, in that intermediates therein may be targets for drug development. Hence, the determinants for the magnitudes of binding rate constants and for binding mechanisms are both of great interest.

Early theoretical developments for binding reactions focused on rigid reactant molecules (72, 77, 101). The simplest such model consists of a spherical protein binding with point-like ligands (Figure 1*a*). The Smoluchowski approach assumes that the ligand molecules are independent. Furthermore, the binding reaction is simplified as irreversible and modeled by an absorbing boundary condition at the surface of the protein. Let $S_N(t)$ be the probability that, up to time t , the binding reaction has not occurred when the protein is surrounded by N ligands (in a volume V). The independence of the ligand molecules leads to

$$S_N(t) = [S_1(t)]^N. \quad 1$$

To obtain a rate equation and derive an expression for the rate constant of the binding reaction, one calculates the time derivative

$$\frac{dS_N(t)}{dt} = N \frac{dS_1(t)}{dt} S_{N-1}(t). \quad 2$$

The survival probability $S_1(t)$ for a single protein–ligand pair can be obtained from a volume integral of the pair distribution function $P(\mathbf{r}, t)$, which represents the relative density of the ligand at position $\mathbf{r}$ and time t , and is given by

$$S_1(t) = \frac{1}{V} \int_V d^3\mathbf{r} P(\mathbf{r}, t). \quad 3$$

By definition, the value of $P(\mathbf{r}, t)$ is 1 at an infinite protein–ligand separation. Taking the thermodynamic limit (i.e., N and $V \rightarrow \infty$ but the ligand concentration N/V remains a constant, denoted by C), we see that Equation 2 resembles a rate equation, namely

$$\frac{dS_N(t)}{dt} = -k_b(t) C S_N(t), \quad 4$$

with a time-dependent rate coefficient

$$k_b(t) = - \int_V d^3\mathbf{r} \frac{\partial P(\mathbf{r}, t)}{\partial t}. \quad 5$$

$P(\mathbf{r}, t)$ can be calculated by solving a diffusion equation governing the protein–ligand relative motion, leading to, for the model in Figure 1a,

$$k_b(t) = 4\pi DR [1 + R/(\pi Dt)^{1/2}], \quad 6$$

where D is the relative diffusion constant and R is the radius of the protein. Experiments usually cannot resolve the transient phase of $k_b(t)$ and only measure the steady-state value (i.e., at $t \rightarrow \infty$), which is referred to as the rate constant and denoted as k_b . The rate constant, $4\pi DR$, of Equation 6 is well known.

Within the restriction of rigid reactant molecules, a number of important factors have been treated theoretically and computationally, including intermolecular interactions, site-specific restraints for binding reactions, and rotational diffusion of reactant molecules (101). For example, when a centrosymmetric interaction potential, $U(r)$, is present, the rate constant becomes

$$k_D(R) = 4\pi D / \int_R^\infty dr [r^2 e^{-\beta U(r)}]^{-1}, \quad 7$$

where r denotes the radial distance and β denotes the inverse of the product of the Boltzmann constant and the absolute temperature (17). In the 1980s, the McCammon group pioneered the use of Brownian dynamics simulations for computing protein–ligand binding rate constants, enabling atomically detailed shape representation for the reactant molecules (55). Subsequently, more and more sophisticated methods have been developed to tackle the challenges, in particular the treatment of intermolecular electrostatic interactions, in modeling the binding kinetics of proteins with macromolecular targets (2, 23, 56).

The possibility of protein conformational fluctuations during the binding process was first considered by McCammon & Northrup (51). They introduced a simple gating model, in which the protein stochastically switches between two conformations, here referred to as inactive and active; only the latter allows the binding reaction to proceed. Szabo et al. (78) realized that the gating model is mathematically equivalent to a rate equation for the conversion between the two conformations (Figure 1b). The rates for the conformational interconversions were assumed to be fixed, regardless of where the ligands are; therefore this model represents a scenario of indifferent switch. By solving for the pair distribution function, now considering both protein–ligand relative diffusion and protein conformational gating, a binding rate coefficient can be obtained (107). This model has been implemented or extended to treat the binding kinetics for a number of actual protein–ligand systems (10, 27, 61, 76, 84, 109). It has been recognized, however, that the binding kinetics of a gating

protein in the presence of N ligands, a many-body problem, may no longer be reduced to a problem involving a single protein–ligand pair (106). That is, a rate equation like Equation 4 may no longer be appropriate for the many-body problem. The reason is that the ligands lose independence because they all see the same conformational switch of the protein and hence all simultaneously respond to the resulting change in boundary condition (from reflecting to absorbing or vice versa; Figure 1*b*).

The indifferent-switch scenario contradicts the current view of the energy landscapes of typical protein–ligand systems. The energy landscapes should favor inactive conformations while the protein and ligand are far apart but favor active conformations when the ligand reaches the binding site (Figure 1*c*). By detailed balance, the change of energy landscape in conformational space means that the ratio of the interconversion rates between the two conformational states varies as the ligand moves from afar to the binding site (8, 100). This scenario, where the conformational interconversion rates are adaptive to the protein–ligand separation, has been referred to as induced switch (102). Analytical and simulation results have been obtained for the protein–ligand binding rate coefficients of model systems by solving the single-pair problem (8, 100). However, the relevance of this rate coefficient to the binding kinetics of the many-body problem (Figure 1*d*) remains to be clarified (see below).

The induced-switch scenario also captures the essence of the interplay between intermolecular and intramolecular motions that gives rise to different binding mechanisms (Figure 1*c*). In a prototypical induced-fit pathway (41), a ligand binds to the protein while it is in the inactive conformation to form a loose complex. Interactions of the ligand with the protein then induce the latter into the active conformation; further tightening of the protein–ligand interactions leads to the native complex. A competing pathway is conformational selection (also known as population shift) (7, 48), whereby the protein, while away from any ligand, makes an excursion into the active conformation. The active conformation is then quickly recognized by a ligand, and formation of the native complex ensues. There has been much debate about whether induced fit or conformational selection is the dominant pathway of protein–ligand binding. Some of this debate undoubtedly can be traced to imprecise definition and the qualitative nature of these mechanistic descriptions. It is now known that pure induced fit and conformational selection occur only in extreme cases, that binding mechanisms usually are mixtures of different pathways, and that the mixing ratios can be quantified (30, 34).

Of course, a two-state model may be too simple for the energy landscapes of many proteins in conformational space. In addition, ligand molecules may also access various conformations, and the accessible conformations may also be different depending on whether they are away from proteins or within binding sites. For more complex systems where multiple pathways and multiple intermediates are involved, simply classifying binding mechanisms as predominantly induced fit or conformational selection is not particularly informative anymore. A case in point is the binding of intrinsically disordered proteins (IDPs) to structured targets. It can be envisioned that the binding process proceeds through multiple pathways, each, for example, starting with the docking of some initial segment of the IDP and ending with the structural coalescence of some final segment on the

target surface (105). The challenge is to establish a framework for modeling the binding process such that both the binding rate constant and the binding mechanism can be determined in quantitative terms.

From the preceding paragraphs, it is clear that great physical insight into binding kinetics can be gained from the following:

1. foundational developments that reduce a rigorous formulation of the many-body kinetic problem to phenomenological rate equations that experimentalists can use to fit data;
2. computational algorithms that enable the calculation of binding rate constants from accurate modeling of intermolecular and intramolecular motions within a single protein–ligand pair;
3. theoretical studies that identify the microscopic properties that dictate binding rates and binding mechanisms; and
4. experimental studies that probe these determinants.

Below, we review the literature in these areas, with a focus on papers published since 2010.

THEORETICAL FOUNDATION

The very concept of binding rate constant is premised on the independence among ligand molecules in the binding kinetics and hence the reduction of the many-body problem to the problem of a single protein–ligand pair. As already noted, when the protein undergoes conformational switches during the binding process, a rate constant calculated from solving the protein–ligand pair problem may not be adequate to describe the many-body binding kinetics. Moreover, even when the solution of the single-pair problem gives a rate constant that adequately capitulates the binding kinetics, it may not be useful for quantitatively determining the binding mechanism. Here, we clarify these issues using a many-body model system for which the binding kinetics has been determined through Brownian dynamics simulations (30).

Many-Body Problem, Four-State Model, and Single-Pair Reduction

The many-body model considered here (Figure 1*d*) was first introduced at the single-pair level (100). The protein is a sphere (with radius R), and the binding site is a thin shell (with radius between R and R_1). The protein can interconvert between an inactive conformation and an active conformation; in the latter case, a ligand inside the binding site can undergo an irreversible binding reaction (modeled as a rate process) to form the native complex. The conformational interconversion rates depend on the locations of the ligands. They have values $\omega_{\infty\pm}$ when all the ligands are far away but satisfy the following detailed-balance relation when the ligands are at positions $\{\mathbf{r}_n\}$:

$$\frac{\omega_+(\{\mathbf{r}_n\})}{\omega_-(\{\mathbf{r}_n\})} = \frac{\omega_{\infty+} e^{-\beta \sum_n U_a(\mathbf{r}_n)}}{\omega_{\infty-} e^{-\beta \sum_n U_i(\mathbf{r}_n)}}, \quad 8$$

where $U_a(\mathbf{r}_n)$ and $U_i(\mathbf{r}_n)$ are the interaction energies of ligand n when the protein is in the active and inactive conformations, respectively.

From Brownian dynamics simulations, the survival probability, $S_N(t)$, for the many-body system was monitored. For a range of model parameters, $S_N(t)$ could be fit to the single exponential

$$S_N(t) = 1 - (1 - S_0)e^{-\mathcal{K}_{\text{app}}t}, \quad 9$$

where S_0 is the (small) amplitude of a fast phase and $\mathcal{K}_{\text{app}}$ is the apparent rate of the binding kinetics. The results for $\mathcal{K}_{\text{app}}$ from the many-body simulations allowed the test of reduced models. In particular, the following four-state rate-equation model has been invoked in many studies (9, 14, 24, 34, 38, 39, 75, 82):

where $P_{\infty i}$ and $P_{\infty a}$ denote the protein in the inactive and active conformations, respectively, while the ligands are far away, whereas $P_i \cdot L$ and $P_a \cdot L$ denote the counterparts when a ligand is in the binding site (forming a loose complex); $P_a \cdot L$ is assumed to proceed immediately to the native complex (Figure 1c). In our case, we assigned the interconversion rates of the unbound protein to be $\omega_{\infty\pm}$, and those of the loosely bound protein to be $\omega_{\pm}$. Furthermore, we assigned the bimolecular rate constants (k_{i+} and k_{a+}) for forming a loose complex, whether the protein was in the inactive or active conformation, to be the Smoluchowski rate constant $4\pi DR_1$. Finally the unbinding rate constants (k_{i-} and k_{a-}) were set on the basis of the equilibrium constants $k_{i+}/k_{i-} = 4\pi(R_1^3 - R^3)\exp(-\beta U_i)/3$ and $k_{a+}/k_{a-} = 4\pi(R_1^3 - R^3)\exp(-\beta U_a)/3$. By making the steady-state approximation for the intermediates $P_{\infty a}$ and $P_i \cdot L$ (valid when these species do not accumulate to a significant extent), the four-state model predicts the following apparent rate of the binding reaction:

$$\mathcal{K}_{\text{app}} = \frac{\omega_{\infty+}k_{a+}C}{\omega_{\infty-} + k_{a+}C} + \frac{\omega_+k_{i+}C}{\omega_- + k_{i-}}. \quad 10$$

In Figure 2a, we compare the predictions of Equation 10 with the simulation results. Equation 10 proves to be quite accurate, except when both the conformational interconversion rates are very low and the ligand concentration is very high [a condition where the many-body nature of the system is most prominent (106)]. Given this encouraging finding, it will be interesting to construct rate-equation models for more realistic protein–ligand systems (including identifying intermediates and determining the rate constants between states) (32) and test how well they recapitulate the many-body binding kinetics.

We now specialize Equation 10 to the limit of low C , where $\mathcal{K}_{\text{app}}$ becomes proportional to C and, correspondingly, the proportionality constant can be identified as the protein–ligand binding rate constant, given by

$$k_b = \frac{\omega_{\infty+}}{\omega_{\infty-}} k_{a+} + \frac{\omega_+}{\omega_+ + k_{i-}} k_{i+}. \quad 11$$

This expression can be taken as accurately representing the many-body simulation results. A binding rate constant can also be determined by solving the single-pair problem (8, 100). The derivation is laborious but straightforward, and the final result for k_b is long, so it is not reproduced here. In Figure 2*b*, we show numerical results of this k_b and compare them to the values given by Equation 11. The close agreement demonstrates that the solution of the single-pair problem yields very accurate results for the binding rate constant. We should point out that whereas the implementation of the four-state model may require improvisation, the single-pair problem is well formulated for any protein–ligand system, and the resulting binding rate constant is unequivocally defined.

In short, the rate constant calculated from solving the protein–ligand pair problem should accurately describe the many-body kinetic problem when the ligand concentration is low and interconversion between protein conformational states is not exceedingly slow. At higher ligand concentrations, phenomenological rate equations may be devised to model the many-body binding kinetics, although what intermediates to introduce (for theorists) or how to interpret them (for experimentalists) may not always be clear or unique. When high ligand concentration is combined with very slow internal dynamics, the many-body nature of the system could be so strong as to preclude the relevance of rate-equation models.

A Quantitative Measure of Binding Mechanism

In both experimental and computational studies, binding mechanisms have mostly been described in qualitative terms. This lack of rigor has caused debate and confusion. In many studies, the detection of active conformations for the unbound protein has been taken as the indicator for conformational selection as the dominant pathway (5, 11, 18, 40, 42, 43, 67, 73, 80, 82, 87, 92). As has been noted (22, 30), every conformation, whether inactive or active, is sampled with an equilibrium probability given by the Boltzmann distribution; whether active conformations can be detected depends on the sensitivity of the probe used and therefore must not decide the binding mechanism.

Hammes et al. (34) introduced a quantitative measure of binding mechanism, based on the four-state model. In this model, the inactive unbound protein (i.e., $P_{\infty i}$) can reach the native complex (i.e., $P_a \cdot L$) via two pathways: The one with the active unbound protein (i.e., $P_{\infty a}$) as the intermediate can be recognized as the conformational-selection pathway, whereas the one with the loosely bound complex (i.e., $P_i \cdot L$) as the intermediate can be recognized as the induced-fit pathway (cf. Figure 1*c*). Hammes et al. realized that, in general, both pathways contribute to the binding process and the binding mechanism inevitably is a mixture. Moreover, the relative contribution of a given pathway is the reactive flux through that

pathway calculated as a fraction of the sum of the reactive fluxes through both pathways. The first and second terms on the right-hand side of Equation 10 can be recognized as the reactive fluxes through the conformational-selection and induced-fit pathways, respectively. Hence, the induced-fit fraction is

$$\Phi_{\text{IF}} = \frac{\omega_+ k_{i+} / (\omega_+ + k_{i-})}{\omega_{\infty+} k_{a+} / (\omega_{\infty-} + k_{a+} C) + \omega_+ k_{i+} / (\omega_+ + k_{i-})} \quad 12$$

The conformational-selection and induced-fit pathways dominate when $\Phi_{\text{IF}} \rightarrow 0$ and 1, respectively. Importantly, this quantitative measure of binding pathways depends on the ligand concentration. That ligand concentration, an extrinsic factor, affects the binding mechanism is still an underappreciated conclusion.

Greives & Zhou (30) proposed a microscopic-level, precise definition of binding events as conformational selection or induced fit for the many-body model of Figure 1*d*. The crucial difference between these two types of binding events is whether the last inactive-to-active transition of the protein before forming the native complex occurs without or with a loosely bound ligand molecule. The many-body simulation results for Φ_{IF} and those predicted by Equation 12 are displayed in Figure 3*a*. The four-state model proves quite accurate also for the induced-fit fraction.

Note that Φ_{IF} increases both when the ligand concentration increases and when the conformational interconversion rates increase. The dependences on both the extrinsic factor (i.e., C) and the intrinsic property (i.e., ω_-) are more clearly illustrated by a two-dimensional surface (Figure 3*b*). Conformational selection dominates only when both the ligand concentration is low and the protein internal dynamics is slow. As already mentioned, the effect of the ligand concentration on Φ_{IF} was first recognized by Hammes et al. (34). The effect of protein internal dynamics on binding mechanism was recognized when the induced-switch scenario was first treated theoretically at the single-pair level (100). That treatment further predicted that with increasing interconversion rates, the binding mechanism continuously shifts from conformational selection to induced fit.

As explained by Greives & Zhou (30), for conformational selection to prevail, the protein must first make an inactive-to-active transition while free of any loosely bound ligand molecule and then stay in the active conformation until a ligand molecule enters the binding site and completes the binding reaction. At a low ligand concentration, the protein is free of any ligand molecule for most of the time, including some occasions when the protein switches to the active conformation, thus allowing the first requirement of conformational selection to be fulfilled. If after such an occasion, a ligand molecule enters the binding site and conformational interconversion is slow, then the protein will stay in the active conformation until the binding reaction occurs, thus fulfilling the second requirement. An increase in ligand concentration makes it more likely to violate the first requirement of conformational selection, whereas an increase in interconversion rates makes it more likely

to violate the second requirement. Either way, the binding mechanism will shift toward induced fit.

The solution of a single-pair problem was important in identifying protein internal dynamics as a determinant of binding mechanism and in predicting the qualitative trend that the binding mechanism shifts toward induced fit in the limit of very fast internal dynamics (100). However, with the reduction of the many-body kinetic problem to a single-pair one, the ability to quantitatively determine the binding mechanism is compromised. This point should be kept in mind when making inferences about binding mechanisms from approaches that are restricted to a single protein–ligand pair, which is the focus of the next section.

ALGORITHMS FOR COMPUTING BINDING RATE CONSTANTS

In principle, the approach of many-body simulations presented in the preceding section on the highly simplified model of Figure 1*d* can be applied to any protein–ligand system to determine the binding rate constant k_b . This would entail simulating the binding process over a range of ligand concentrations, fitting the many-body survival probability $S_M(t)$ to an exponential function to obtain the apparent binding rate, and, finally, finding k_b as the slope in the dependence of the apparent binding rate on the ligand concentration. This procedure is similar to how binding rate constants are typically measured by experimentalists. In a heroic study, Shan et al. (70) carried out such brute-force many-body simulations, with the protein, ligand molecules, and solvent molecules all represented at the all-atom level, on a special-purpose computer, Anton. With this realistic representation of the protein–ligand system, observing spontaneous binding of the ligand to correctly form the native complex was a significant achievement. However, because the authors were able to observe only a very small number of binding events (one event for one ligand and three events for another ligand) during the simulations, they could obtain only order-of-magnitude estimates for the binding rate constants. The fact that the brute-force many-body simulation approach falls short of a precise calculation of the binding rate constant, even with arguably the best effort, points to the need for algorithms specifically designed for this purpose.

Below, we outline the basic ideas behind a number of algorithms for calculating binding rate constants [Figure 4; for a related review with emphasis on protein–drug binding, see Romanowska et al. (66); also see the article by Zuckerman & Chong (111) in this volume]. As already emphasized, the concept of a rate constant itself implies the validity of the single protein–ligand pair description. Some of the algorithms directly start from this description; others are based on treating the binding kinetics as a pseudounimolecular reaction but with the theoretical underpinning not fully clear.

Brownian Dynamics Simulations of Rigid Protein and Ligand Molecules

The McCammon group opened the field of computing binding rate constants for atomistic proteins and ligands by implementing the Smoluchowski approach through Brownian dynamics simulations of ligand diffusion (55). In the McCammon group's algorithm (hereafter referred to by the year of its publication, 1984), two concentric spherical surfaces enclosing the protein molecule are introduced—one (at radial distance $r = b$) for starting Brownian trajectories, and the other (at $r = q$) for terminating trajectories that run away from

the protein (Figure 4a). The binding reaction is modeled by an absorbing boundary condition imposed over the binding site; the rest of the protein surface is reflecting. Each trajectory terminates either at the binding site or on the q surface. By launching many trajectories at $r = b$, one obtains the probability, $p(b)$, that these trajectories terminate at the binding site rather than on the q surface. The binding rate constant is

$$k_b = \frac{k_D(b)p(b)}{1 - [1 - p(b)]k_D(b)/k_D(q)}, \quad 13$$

where $k_D(b)$ denotes the expression in Equation 7. When b is large and the binding site is deeply buried, the value of $p(b)$ can be very small, such that an excessively large number of trajectories have to be launched to ensure its precise determination. Equation 13 was originally derived using branching diagrams but was later derived analytically, under the condition that the interaction potential is centrosymmetric at $r > b$ and the pair distribution function is centrosymmetric at $r > q$ (96). An intermediate result of that derivation is

$$k_b = k_D(b)\eta(b), \quad 14$$

where $\eta(b)$ denotes the capture probability (i.e., the probability that Brownian trajectories started at $r = b$ terminate at the binding site rather than escape to infinity).

The q surface was introduced as a way to determine $\eta(b)$ indirectly, because any finite Brownian trajectory can never tell whether the ligand has escaped to infinity. The need for the q surface can be eliminated by an analytical treatment, whereby a ligand outside the b surface is, according to a known probability ratio, either designated as having escaped to infinity or placed back on the b surface with a known probability distribution (47). The latter is an example of the first-hitting distribution. Recently, Held et al. (35) reformulated the problem in discretized space, thereby transforming ligand diffusion into a Markov jump process. The solution for the capture probability, also known as the committor, was found according to the transition path theory (20, 52, 54). The 1984 algorithm has been extended to model protein–protein association kinetics, accounting for both translational and rotational diffusion of the molecules and approximately treating the complex intermolecular electrostatic interactions along the Brownian trajectories (23, 56, 94).

An alternative to the 1984 algorithm was proposed in 1990 (95) (Figure 4b). Here, the binding reaction is modeled as a rate process (with rate γ) inside the binding site. The time-dependent rate coefficient is obtained as

$$k_b(t) = k_b(0) \langle S(t | \mathbf{r}_0) \rangle^*, \quad 15$$

where $S(\mathbf{r}_0)$ is the probability that after starting the Brownian trajectory at a position $\mathbf{r}_0$ within the binding site (signified by a superscript asterisk), the binding reaction has not occurred at time t , and $\langle \dots \rangle^*$ means averaging over all initial positions within the binding site. The initial value of rate coefficient is

$$k_b(0) = \int^* d^3\mathbf{r}_0 \gamma e^{-\beta U(\mathbf{r}_0)} = V^* \gamma \langle e^{-\beta U(\mathbf{r}_0)} \rangle^*, \quad 16$$

where $U(\mathbf{r})$ denotes the intermolecular interaction potential and V^* denotes the volume of the binding site. For long-ranged electrostatic interaction potentials, Brownian dynamics simulations following the 1990 algorithm showed that $\langle S(\mathbf{r}_0) \rangle^*$ is insensitive to the presence of the potentials (97). This observation hinted that the effect of long-range electrostatic interactions on the rate coefficient can be approximated by the effect on the initial value $k_b(0)$ and is therefore captured by the average Boltzmann factor $\langle e^{-\beta U(\mathbf{r}_0)} \rangle^*$. Subsequent analytical results further clarified that the validity of this approximation requires, in addition to the interaction potential being long ranged, the binding site being stereospecific (98, 99). We can now express the effect of long-range electrostatic interactions on the binding rate constant by the following formula:

$$k_b = k_{b0} e^{-\beta G_{\text{el}}^*}, \quad 17$$

where k_{b0} is the basal rate constant (i.e., the rate constant when long-range electrostatic interactions are turned off) and G_{el}^* is the electrostatic free energy within the binding site, converted from the average Boltzmann factor. Note that Equation 17 allows us to completely avoid the problem of treating the complex intermolecular electrostatic interactions along Brownian trajectories of protein–protein association (2). Instead, we generate Brownian trajectories while turning off electrostatic interactions to obtain k_{b0} . We then capture the effects of electrostatic interactions by calculating the electrostatic interaction potential within the binding site to obtain G_{el}^* .

Algorithms based on Brownian dynamics simulations of translational and rotational diffusion have the intrinsic limitation that the location where the binding reaction occurs has to be externally specified. The magnitude of the calculated rate constant is very sensitive to the precise position and size of the binding site; therefore, the uncertainty in its specification significantly impairs the predictive power of the algorithms. In a step toward self-containment, Alsallaq & Zhou (2) proposed the rim of the bound-state free energy well in the six-dimensional space of relative translation and rotation as the appropriate location for diffusion-limited protein–protein association. This rim defines an intermediate called the transient complex, in which the two proteins have near-native separation and relative orientation but have yet to form most of the stereospecific native contacts. The computational method combining the 1990 algorithm with the concept of the transient complex has been automated and developed into a web server called TransComp (<http://pipe.sc.fsu.edu/transcomp/>) (63).

Hybrid Simulations of Rigid and Flexible Molecules

When the Smoluchowski approach is implemented through Brownian dynamics simulations of ligand diffusion, the protein molecule does not need to stay rigid. Its motions can be modeled so that their effect on k_b is accounted for. This was done by Wade et al. (86), who supplemented the 1984 algorithm for k_b calculation by also simulating the motions of a flexible loop near the binding site. The loop was modeled as diffusive beads on a string. It is predicted theoretically that the effect of gating motions that control the ligand access to the binding site depends on their timescale relative to the timescale of ligand diffusion (109). When gating motions are fast, they do not hinder the ligand binding rate; the latter was precisely the observation of Wade et al.

From a technical standpoint, simulating protein motions along the Brownian trajectories of the ligand is extremely inefficient because the ligand is away from the protein most of the time (particularly in the 1984 algorithm) and then is not affected by the protein motions. Luty and coworkers (45, 46) recognized this problem and devised an elegant solution by introducing two dividing surfaces (numbered 2 and 3) enclosing the binding site (Figure 4c). Instead of obtaining the capture probability $\eta(b)$ directly as in the 1984 algorithm, one first obtains a series of transition probabilities between the dividing surfaces and the two end states (i.e., the binding site and infinity), to construct a Markov chain. Formally, such a Markov chain is called an absorbing Markov chain for the fact that the end states are traps (i.e., transition probabilities $p_{jj} = 1$); correspondingly, the end states are absorbing states (denoted by index j), and the dividing surfaces are transient states (denoted by index i) (31). The transition probability matrix is

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ p_{21} & 0 & 1-p_{21} & 0 \\ 0 & p_{32} & 0 & 1-p_{32} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad 18$$

In determining the transition probabilities from simulations, the ligand is confined to a specific region: inside dividing surface 3 for p_{21} and outside dividing surface 2 for p_{32} . One can choose to make the protein flexible in the first case and rigid in the second case. Moreover, in the first case, instead of implicit-solvent Brownian dynamics simulations, one can use explicit-solvent molecular dynamics simulations to more realistically model the conformational dynamics of the protein when the ligand is loosely bound. Also importantly, because the binding site and the dividing surfaces are relatively closely spaced, the values of p_{21} and p_{32} are expected to be much greater than $\eta(b)$, and hence the number of trajectories required is also reduced.

The probabilities, η_i , that the chain starts at transient states i and finally ends at absorbing state 1 are obtained by multiplying $\mathbf{P}$ with itself an infinite number of times:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ \eta_2 & 0 & 0 & 1-\eta_2 \\ \eta_3 & 0 & 0 & 1-\eta_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \lim_{l \rightarrow \infty} \mathbf{P}^l \equiv \mathbf{P}^\infty. \quad 19$$

Alternatively, η_i can be determined with the help of the so-called harmonic function $\mathbf{h}$ for $\mathbf{P}$ (31), defined by

$$\mathbf{h} = \mathbf{P} \cdot \mathbf{h}. \quad 20$$

By repeatedly replacing $\mathbf{h}$ in the right-hand side by $\mathbf{P} \cdot \mathbf{h}$, one finds that Equation 20 holds when $\mathbf{P}$ is replaced by $\mathbf{P}^\infty$. The latter result allows us to express η_i in terms of $\mathbf{h}$ as

$$\eta_i = \frac{h_i - h_4}{h_1 - h_4}. \quad 21$$

Using Equation 20 to solve for h_i and inserting the results in Equation 21, we find

$$\eta_2 = \frac{p_{21}}{1 - (1 - p_{21})p_{32}} \quad 22$$

and

$$\eta_3 = \eta_2 p_{32}. \quad 23$$

Had we chosen the b and q surfaces as the two transient states, p_{21} and η_2 would be $p(b)$ and $\eta(b)$, respectively, and Equation 22 expressing $\eta(b)$ in terms of $p(b)$ when inserted into Equation 14 would lead to Equation 13. If we choose surface 2 and the b surface as the two transient states, then Equation 23 becomes

$$\eta(b) = \eta_2 \eta(b \rightarrow 2), \quad 24$$

where $\eta(b \rightarrow 2)$ is the capture probability for starting on the b surface and ending on dividing surface 2. Using the last result in Equation 14, we find the binding rate constant as

$$k_b = k_2 \eta_2 \quad 25$$

with

$$k_2 = k_D(b)\eta(b \rightarrow 2). \quad 26$$

By comparing Equations 14 and 26, we can recognize k_2 as the rate constant when dividing surface 2 is the binding site. Equation 25 can then be viewed as a generalization of Equation 14, in which the spherical b surface is replaced by dividing surface 2.

In short, the algorithm of Luty et al. entails determining p_{21} from molecular dynamics simulations and p_{32} and $\eta(b \rightarrow 2)$ from Brownian dynamics simulations. Recently, this algorithm was revamped (83), using the first-hitting distribution, as suggested by the milestoning theory (49; see below), for launching molecular and Brownian dynamics simulations. Below, we will contrast the use of the first-hitting distribution with other choices when calculating η_2 for use in Equation 25.

A different method for treating protein flexibility was developed by Greives & Zhou (29) (Figure 4*d*). The aim is to break the full problem of k_b calculation into two separate problems: an exterior problem, where the ligand is confined outside a dividing surface S^* (to be signified by a superscript asterisk) and the protein is kept rigid; and an interior problem, where the ligand is confined within the dividing surface and the protein is fully flexible. The starting point of this BDflex method is the following expression for the rate constant (102):

$$k_b = - \int^* ds \mathbf{n} \cdot \mathbf{J}_E(\mathbf{r}) \eta(\mathbf{r}), \quad 27$$

where $\mathbf{n}$ is a unit vector along the outward normal direction of the dividing surface, $\mathbf{J}_E(\mathbf{r})$ is the flux density of the exterior problem, where an absorbing boundary condition is imposed on the dividing surface, and $\eta(\mathbf{r})$ is the capture probability starting at position $\mathbf{r}$ on the dividing surface.

Note that the rate constant of the exterior problem (i.e., the rate constant for being absorbing on S^*) is

$$k_E = - \int^* ds \mathbf{n} \cdot \mathbf{J}_E(\mathbf{r}). \quad 28$$

If S^* is chosen such that $\eta(\mathbf{r})$ is uniform over it (a surface satisfying this requirement is called an isocommittor), then Equation 27 reduces to

$$k_b = k_E \eta_{IC}^*, \quad 29$$

where η_{IC}^* denotes the uniform value of $\eta(\mathbf{r})$ on S^* . If S^* is not an isocommittor surface, we may replace η_{IC}^* by the average

$$\langle \eta \rangle^* = \int^* ds \eta(\mathbf{r}) / \int^* ds, \quad 30$$

but the resulting k_b is only approximate. However, the exact result for k_b is obtained if the average over the nonisocommittor surface uses $-\mathbf{n} \cdot \mathbf{J}_E(\mathbf{r})$, the first-hitting distribution of the exterior problem, as the weighting factor, yielding

$$\eta_{FH}^* = \int^s ds \mathbf{n} \cdot \mathbf{J}_E(\mathbf{r}) \eta(\mathbf{r}) / \int^* ds \mathbf{n} \cdot \mathbf{J}_E(\mathbf{r}). \quad 31$$

Then,

$$k_b = k_E \eta_{FH}^*. \quad 32$$

We can now recognize Equation 25 as an example (where $S^* = \text{surface 2}$) of Equations 29 or 33 and conclude that it is exact if surface 2 is an isocommittor or η_2 represents the average weighted by the first-hitting distribution. If surface 2 is not an isocommittor and the average is not weighted (as in Equation 30), then Equation 25 is only approximate.

To define an interior problem that complements the exterior problem and leads to the determination of k_b , we need a boundary condition for $\eta(\mathbf{r})$ that is imposed on S^* (as the outer boundary of the interior problem). To that end, we use another expression for the rate constant,

$$k_b = \int^* ds \mathbf{n} \cdot \mathbf{J}(\mathbf{r}) = \int^* ds e^{-\beta U(\mathbf{r})} D \partial \eta(\mathbf{r}) / \partial n, \quad 33$$

where $\mathbf{J}(\mathbf{r})$ is the flux density of the full problem. From here on, we assume that both $\eta(\mathbf{r})$ and its normal derivative $\partial \eta(\mathbf{r}) / \partial n$ are approximately uniform over S^* , which is valid either when S^* is close to being an isocommittor or when the pair distribution function stays nearly equilibrated among all the positions on S^* . Then, comparing the right-hand sides of Equations 27 and 33, we obtain

$$D \partial \eta(\mathbf{r}) / \partial n \int^* ds e^{-\beta U(\mathbf{r})} \approx -\eta(\mathbf{r}) \int^* ds \mathbf{n} \cdot \mathbf{J}_E(\mathbf{r}), \mathbf{r} \in S^* \quad 34$$

or

$$D\partial\eta(\mathbf{r})/\partial n \approx \kappa_E \eta(\mathbf{r}), \mathbf{r} \in S^*, \quad 35$$

where

$$\kappa_E = \frac{k_E}{\int^* ds e^{-\beta U(\mathbf{r})}}. \quad 36$$

Equation 35 is the desired boundary condition; it is a partially absorbing boundary condition, with the strength of absorption, κ_E , determined by the solution of the exterior problem.

We now see that k_b can be obtained by separately solving first the exterior problem and then the interior problem. The solution of the exterior problem yields k_E and the first-hitting distribution. k_E is used for specifying the partially absorbing boundary condition on the outer surface for the interior problem, whereas the first-hitting distribution can be used for launching ligand trajectories. The solution of the interior problem yields η_{FH}^* (or η_{IC}^* if an isocommittor surface is chosen as S^*). Finally, the product of k_E and η_{FH}^* gives k_b . Whereas the exterior problem is solved through Brownian dynamics simulations with the protein kept rigid, the interior problem can be solved by molecular dynamics simulations to allow for realistic modeling of protein motions.

Molecular Dynamics Simulations of Fully Flexible Molecules

Recently, explicit-solvent molecular dynamics simulations have been used to fully determine protein–ligand binding rate constants. Buch et al. (6) carried out 495 replicate 100-ns simulations. By constraining the ligand to a small volume around the native binding site, they observed 187 binding events in these relatively short simulations. They then constructed Markov state models (MSMs) to identify binding pathways and calculate the binding rate constant. MSMs are a way to extract longtime kinetic information from multiple short simulations by using a discrete-state stochastic description for the kinetic behavior of the system (12). This is similar to the utility of the absorbing Markov chain presented above, whereby the probabilities for reaching end states are predicted from transition probabilities between closely spaced intermediates (more of the relation between MSMs and Markov chains given below). To construct an MSM, one first partitions the conformational space of the system into a set of discrete states (denoted by index i or j). Each conformation sampled from the simulations is assigned a state index. The second ingredient of the MSM is the transition probability matrix $\mathbf{P}$, whose elements p_{ij} are the probabilities that, if at any time the system is in state i , then, after a fixed lag time τ , the system is found in state j . MSMs also assume that the system is ergodic (i.e., it can go from every state to every other state). Ergodicity ensures that the system can reach equilibrium at infinite times. To satisfy the ergodicity requirement, sometimes p_{ij} are defaulted to a small value when transitions from state i to state j are not sampled in the simulations or analysis is restricted to the largest ergodic subset.

Mathematicians formally refer to MSMs as ergodic Markov chains and have developed an extensive theory for them (31). In particular, the powers of $\mathbf{P}$ approach a limiting matrix

$$\mathbf{W} \equiv \lim_{l \rightarrow \infty} \mathbf{P}^l \quad 37$$

that has the same row vector, $\mathbf{w}$. The latter can be directly obtained from solving the equation

$$\mathbf{w} = \mathbf{w} \cdot \mathbf{P}. \quad 38$$

The components of $\mathbf{w}$ are the equilibrium probabilities of the individual states, which can be converted into the potential of mean force by the Boltzmann relation. The mean first passage time, m_{ij} (i.e., the number of individual transitions between states that the system, starting in state i , has to go through to reach state j for the first time) can be determined from the fundamental matrix

$$\mathbf{Z} \equiv (\mathbf{I} - \mathbf{P} + \mathbf{W})^{-1}, \quad 39$$

where $\mathbf{I}$ is the identity matrix. Specifically,

$$m_{ij} = \frac{z_{jj} - z_{ij}}{w_j}. \quad 40$$

Because each transition takes a fixed time τ , in actual time units, the mean first passage time is $m_{ij} \tau$.

Buch et al. (6) constructed MSMs at different levels of coarse graining. In one, the partition into states was according to a rectangular grid for the projected ligand position on a plane through the native binding site. The potential of mean force shows minima at the native binding site and three intermediate sites. Further lumping into five states (including the unbound state) using a Voronoi tessellation allowed the authors to identify preferred pathways from the unbound state to the native binding site. The authors also partitioned states according to a parallelepipedic grid and used the resulting potential of mean force for final lumping into just two states, bound and unbound. After calculating the mean first passage time, MFPT_{ub} , from the unbound state to the bound state, they obtained the binding rate constant as

$$k_b = \frac{1}{\text{MFPT}_{\text{ub}} C}, \quad 41$$

where C is the concentration of the ligand (one molecule in the restrained volume). The simulations have been extended by Plattner & Noe (62), who, after constructing new MSMs, identified up to seven parallel but linked binding pathways. The unbiased simulations of Buch et al. were also combined very recently with biased (i.e., umbrella sampling) simulations to make more robust estimates of binding and unbinding rate constants, using transition-based reweighting analysis (91).

Yu et al. (93) presented a method for k_b calculation based on the milestoning theory. This theory closely mirrors the MSM approach and uses hypersurfaces, referred to as milestones, in the conformational space of the system as intermediate states between the end states of a reaction (21). Yu et al.'s implementation used isocommittor surfaces as milestones (81). More specifically, the milestones were defined in the space of collective variables, and the isocommittor surfaces were approximated by hyperplanes that are locally orthogonal to a minimum free energy path between the end states (50). For the protein–ligand binding problem, Yu et al. added a spherical surface enclosing the protein and divided that surface into milestones that all faced the exterior. Fluxes through interior milestones were determined from molecular dynamics simulations, but the inward fluxes through the exterior-facing milestones were estimated by proposing a constant flux density, $4\pi DRC/4\pi R^2$, which was obtained by treating the protein as a spherical absorber with radius R . Finally, the inverse of the mean first passage time from a collection of exterior-facing milestones to the milestone facing the binding site was taken as the pseudounimolecular rate constant for binding through that particular pathway, and the corresponding binding rate constant was obtained upon dividing by the ligand concentration C , just as stated by Equation 41.

Yu et al.'s proposed constant flux density into exterior-facing milestones is open to criticism. It is intriguing to ask whether this potential shortcoming can be remedied by introducing the solution of the exterior problem from BDflex.

DETERMINANTS FOR THE MAGNITUDES OF BINDING RATE CONSTANTS

In principle, the magnitude of the binding rate constant is dictated by the energetics and dynamics of the intra- and intermolecular degrees of freedom in the protein–ligand pair. Identification of single factors that can tune the magnitudes of k_b by orders of magnitude not only contributes to fundamental knowledge but also can be the basis for molecular design.

Long-Range Electrostatic Attraction

Our best understanding on the determinants of k_b magnitudes is for binding processes that are rate limited by translational and rotational diffusion. In this case, either the reactant molecules have little flexibility or the internal dynamics of the molecules occur on timescales much faster than the timescales of translational and rotational diffusion so that

the molecules can be effectively treated as rigid. The magnitude of a diffusion-limited k_b depends on the size of the rim of the bound-state free energy well in the six-dimensional space of relative translation and rotation, and on any long-range electrostatic interaction potential between the molecules. For protein–protein association, this rim defines the transient complex, and Equation 17 allows the effects of the rim size and the electrostatic interaction potential to be separated (2, 63). Specifically, the effect of the rim size is contained in the basal rate constant k_{b0} , whereas the effect of the electrostatic interaction potential is fully captured by the electrostatic interaction free energy, G_{el}^* , of the transient complex. The latter appears in the exponent, so changes in G_{el}^* by a few kilocalories per mole correspond to changes in k_b by several orders of magnitude.

Although long-range electrostatic attraction has long been recognized as an important factor for binding rate enhancement (23, 56, 69), the isolation in an exponential as in Equation 17 both enormously simplifies the prediction of its effects and makes it much easier to understand these effects. For example, four four-helix bundle cytokines bind to their receptors with rate constants spanning a 5,000-fold range (4, 16, 37, 88). Calculations using Equation 17 showed that the basal rate constants of the four cytokines differ by no more than threefold, and the vast differences in their receptor-binding rate constants arise mostly from the differences in electrostatic rate enhancement (57). Because of the varying extents of charge complementarity across the protein–protein interfaces, G_{el}^* values differ by as much as 4.1 kcal/mol. Similarly, 1,000-fold variations in k_b for G-actin binding to seven actin-binding proteins can be attributed to disparate electrostatic contributions (59). Using the TransComp web server on 132 nonhomologous protein–protein complexes, the values of the basal rate constants were found to be mostly confined in the range of 10^4 to 10^6 M⁻¹s⁻¹, whereas the electrostatic contributions varied by six orders of magnitude (63). In the case of a ribotoxic enzyme binding to a ribosomal RNA loop, a small displacement of the ribotoxin by a neighboring ribosomal protein to a site with higher charge complementarity leads to a 3.2 kcal/mol decrease in G_{el}^* , which translates into an approximately 200-fold increase in k_b (64).

In comparison, small-molecule ligands cannot carry as many charges as a protein; therefore, the extent to which long-range electrostatic interactions can modulate k_b is not expected to be as dramatic. Still, for charged ligands (e.g., superoxide, acetylcholine, and phosphate), significant electrostatic rate enhancements have been found (25, 35, 36, 45, 71, 79, 83, 85, 90, 104, 109, 110).

If long-range electrostatic attraction can produce enormous rate enhancement, what about short-range attraction? Theoretical calculations show that the latter's effects on k_b are significant only if it is present over very extended surfaces, such as the entire membrane surface in the case of binding to a membrane-bound receptor or the entire surface of a long DNA molecule in the case of binding to a native site therein (108). In these cases, the short-range attraction can keep the ligand loosely bound to the extended surface for long periods of time, during which the search for the native binding site occurs in the space with a reduced dimension. The chance of finding the native site increases with the reduced dimensionality (1), hence the enhancement in k_b .

Gates, Intermediate Sites, and Desolvation

After the ligand diffuses to the protein and binds loosely, either near the native site or at other intermediates sites, the dynamics of the molecules can become rate limiting for forming the native complex. This is easily understood in the case where access to the native site is through a gate (75, 86, 93, 109)—if the rate of gate opening is decreased further and further, eventually it will limit the rate constant of the overall binding process. Although there are many ways to manipulate the magnitude of electrostatic attraction (e.g., mutation, change of salt concentration or pH), it seems difficult to experimentally engineer a change in the rate of gate opening without affecting other properties of the protein.

If the ligand visits intermediate sites before reaching the native site, then the rates of transition among the intermediate and native sites can affect k_b . For example, Buch et al. (6) identified three intermediate sites for their system. Shan et al. (70) in their simulations also identified several intermediate sites, including some that were known sites for binding other ligands, a site that was too shallow in the crystal structure to be recognized previously, and a site separated from the native site by a shell of water. The transition from the last intermediate site to the native site involved desolvation of the ligand, which Shan et al. suggested represented a kinetic barrier.

In general, k_b increases with increasing rates of gate opening, transition among intermediate and native sites, and desolvation. Beyond that, we have relatively little quantitative knowledge on how much the magnitudes of k_b are affected by the rates of these steps within the loosely bound complex. Theoretical models and simulation protocols designed to study these effects and gain better understanding will be very welcome. A difficulty, as just alluded to, is to experimentally test this understanding.

Unbinding rate constants have been absent in this review. Here, we make two comments. First, the binding process comprises steps up to reaching the rim of the bound-state well, whereas the unbinding process includes the escape from the bound-state well. Therefore, the unbinding rate constant, like the binding equilibrium constant, is sensitive to changes in the depth of the bound-state well. Significant such changes can be easily brought by mutations in the native binding site. In contrast, determinants of binding rate constants are not expected to be as sensitive to mutations, even including the ones perturbing long-range attraction in the diffusion-limited case. Second, our present inattention notwithstanding, the unbinding rate constant is widely thought to be a determining factor for drug efficacy, with slow unbinding a desirable property (44, 68).

DETERMINATION AND DETERMINANTS OF BINDING MECHANISMS

Instead of qualitative descriptions, binding mechanisms can and should be quantitatively determined. The underlying idea is that when the protein and ligand can reach the native complex through alternative pathways, each pathway contributes to the total reactive flux. Its contribution as a fraction of the total reactive flux thus stands as a quantitative measure for the relative importance of the given pathway. In general, binding mechanisms are mixtures of different pathways, although they can be dominated by a single pathway. It is worth emphasizing that binding mechanisms depend on the ligand concentration (15, 30, 34);

therefore, the dominant pathway can change when the ligand concentration is changed. In particular, with increasing ligand concentration, a minor pathway (e.g., induced-fit) may become dominant, whereas the dominant pathway (e.g., conformational-selection) may become minor. The quantitative measure using fractional contribution to the total reactive flux is premised on the pathways being parallel (i.e., mutually exclusive). The decomposition into parallel pathways may prove difficult. Below, we summarize existing and desired efforts toward quantifying binding mechanisms.

Computational Determination of Binding Mechanisms

As noted above, methods designed to calculate k_b efficiently usually are based on simulations of a single protein–ligand pair. The single pair can be viewed as the $C \rightarrow 0$ limit of the many-body system. It is therefore of interest to see whether the simulations designed for calculating k_b can also be used for quantifying the binding mechanism in the $C \rightarrow 0$ limit. For the four-state model, by comparing Equations 10 and 11, we can recognize the two terms in the expression for k_b as contributions from the conformational-selection and induced-fit pathways, respectively, namely

$$k_b = k_{b;CS} + k_{b;IF}. \quad 42$$

The induced-fit fraction in the $C \rightarrow 0$ limit can then be obtained as

$$\Phi_{IF} = \frac{k_{b;IF}}{k_b}. \quad 43$$

For the model of Figure 1*d*, the conformational-selection component $k_{b;CS}$ can be obtained from solving the single-pair problem with the ligand diffusion constant set to zero whenever the protein is in the inactive conformation (100) (such that the ligand first waits for the protein to switch to the active conformation and then binds to the protein to form the native complex).

In general, if the binding process can proceed on multiple parallel pathways, then

$$k_b = k_{b;1} + k_{b;2} + \cdots, \quad 44$$

where each term on the right-hand side denotes the rate constant from binding via a particular pathway exclusively. Potentially, intermediate sites such as those identified by Buch et al. (6) in their simulations could be the basis of introducing binding pathways, but because the ligand can transition from any intermediate site to another intermediate site, these pathways are more likely to be nested than parallel (62). The situation is better in the system studied by Yu et al. (93), which features a buried site connected to the bulk solvent through several channels. Pathways through these channels are naturally independent, and

by comparing their relative contributions to the binding rate constant, Yu et al. identified a dominant channel.

Still, the effect of ligand concentration on binding mechanism is missing. This problem was recently addressed by Gu et al. (32). Using MSMs constructed from molecular dynamics simulations, they proposed an extension of the four-state model, in which the inactive state comprised four substates. The latter were treated as independent; therefore, the extended model was a superposition of four separate four-state models. Although this is a very welcome development, the nature of its approximation still needs to be clarified, by, for example, testing against brute-force many-body simulations of the binding kinetics (70).

Experimental Determination of Binding Mechanisms

In the four-state model, the apparent rates for the conformational-selection pathway and for the induced-fit pathways have disparate dependences on ligand concentration (e.g., comparing the two terms on the right-hand side of Equation 10). This disparity has been proposed as the basis for distinguishing binding mechanisms as either conformational selection or induced fit (26, 53, 60, 82). In reality, binding mechanisms are always a mixture, and the mixing ratio is concentration dependent, with induced fit dominating at high concentration (15, 30, 34). Rather than fitting kinetic data to a single branch, experimentalists should fit their data to the full four-state model, although then the fitting parameters may be too numerous to be adequately determined.

Fortunately, several experimental techniques, including nuclear magnetic resonance spectroscopy and single-molecule fluorescence resonance energy transfer (smFRET), enable the direct determination of some of these parameters—namely the rates of interconversion between active and inactive conformations of proteins, with or without bound ligands (9, 13, 39, 74, 89, 92). In a very exciting recent study (39), three-color smFRET allowed simultaneous detection of protein conformational states and ligand binding events, directly showing that ligand binding occurred mostly when the maltose binding protein was in the inactive conformation and hence predominantly proceeded through the induced-fit pathway.

Determinants of Binding Mechanisms

Several determinants of binding mechanisms have been identified. Hammes et al. (34) first recognized ligand concentration as an extrinsic determinant, which was further emphasized by Daniels et al. (15) in an experimental study and clarified by Greives & Zhou (30) through many-body simulations of a model system. Zhou (100) first identified protein conformational interconversion rates as a determinant, which was further clarified in the many-body simulations (30). Recently, Daniels et al. (14) highlighted the effects on the binding mechanism by two related parameters: the ratio between the conformational equilibrium constants of the bound and unbound protein and the corresponding ratio of the inactive-to-active rates. Note that these two parameters are unchanged if all the rates increase or decrease by a common factor. A uniform increase in the rates is predicted to shift the binding mechanism toward the induced-fit pathway (30, 100). Two mutants of the maltose binding protein studied by Kim et al. (39) differ in inactive-to-active and active-to-inactive rates both by approximately tenfold, and thus present an opportunity to test the

latter prediction. Other studies have already shown that binding mechanisms can be shifted by mutation (28) and Mg^{2+} (74).

Beyond Conformational Selection and Induced Fit

When multiple intermediate binding sites exist (see above), simply classifying binding mechanisms as conformational selection or induced fit is no longer particularly informative. An extreme case is IDP binding to structured targets (19). An IDP usually adopts a very extended conformation on its structured target, with stabilization provided less by intramolecular and more by intermolecular interactions. Simultaneous formation of all these latter interactions upon approaching the target is unlikely (thereby precluding the classical conformation-selection pathway as dominant); an attractive alternative is sequential formation of the bound conformation (63, 103). Specifically, in a proposed mechanism called dock and coalesce, some segment of the IDP docks to its subsite and primes the remaining segments for coalescence around their subsites. There are still multiple parallel pathways, as any number of segments of the IDP can serve as the docking segment. These pathways all contribute to the binding rate constant (see Equation 44). If docking by a particular segment yields a much higher rate constant than other alternatives, then that pathway becomes dominant. Using the TransComp web server to predict the rate constants for the docking of various segments and assuming that the coalescing step is not rate limiting, the dominant dock-and-coalesce pathways have been identified for the binding of a number of IDPs, all supported by experimental data (58, 59, 63, 105). Rogers et al. (65) recently presented a scenario of dock and coalesce, with the docking and coalescing segments identified by experimental data for the effects of point mutations within the binding site on the binding rate constant. Interestingly, in addition to direct interactions within the binding site, allosteric effects by distally bound domains may also facilitate the dock-and-coalesce mechanism (33).

Acknowledgments

This research has been supported in part by the National Institutes of Health through grants GM058187, GM088187, and GM118091.

LITERATURE CITED

1. Adam, G., Delbruck, M. Reduction of dimensionality in biological diffusion processes. In: Davidson, N., editor. *Structural Chemistry and Molecular Biology*. San Francisco: W.H. Freeman; 1968. p. 198-215.
2. Alsallaq R, Zhou H-X. Electrostatic rate enhancement and transient complex of protein-protein association. *Proteins*. 2008; 71:320–35. [PubMed: 17932929]
3. Berezhkovskii AM, Szabo A, Greives N, Zhou H-X. Multidimensional reaction rate theory with anisotropic diffusion. *J Chem Phys*. 2014; 141:204106. [PubMed: 25429932]
4. Bernat B, Sun M, Dwyer M, Feldkamp M, Kossiakoff AA. Dissecting the binding energy epitope of a high-affinity variant of human growth hormone: cooperative and additive effects from combining mutations from independently selected phage display mutagenesis libraries. *Biochemistry*. 2004; 43:6076–84. [PubMed: 15147191]
5. Bohnuud T, Kozakov D, Vajda S. Evidence of conformational selection driving the formation of ligand binding sites in protein-protein interfaces. *PLOS Comput Biol*. 2014; 10:e1003872. [PubMed: 25275445]

6. Buch I, Giorgino T, De Fabritiis G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *PNAS*. 2011; 108:10184–89. [PubMed: 21646537]
7. Burgen AS. Conformational changes and drug action. *Fed Proc*. 1981; 40:2723–28. [PubMed: 7297703]
8. Cai L, Zhou H-X. Theory and simulation on the kinetics of protein–ligand binding coupled to conformational change. *J Chem Phys*. 2011; 134:105101. [PubMed: 21405192]
9. Chakrabarti KS, Agafonov RV, Pontiggia F, Otten R, Higgins MK, et al. Conformational selection in a protein-protein interaction revealed by dynamic pathway analysis. *Cell Rep*. 2016; 14:32–42. [PubMed: 26725117]
10. Chang C-E, Shen T, Trylska J, Tozzini V, McCammon JA. Gated binding of ligands to HIV-1 protease: Brownian dynamics simulations in a coarse-grained model. *Biophys J*. 2006; 90:3880–85. [PubMed: 16533835]
11. Changeux J-P, Edelstein S. Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol Rep*. 2011; 3:19. [PubMed: 21941598]
12. Chodera JD, Noe F. Markov state models of biomolecular conformational dynamics. *Curr Opin Struct Biol*. 2014; 25:135–44. [PubMed: 24836551]
13. Clore GM. Interplay between conformational selection and induced fit in multidomain protein–ligand binding probed by paramagnetic relaxation enhancement. *Biophys Chem*. 2014; 186:3–12. [PubMed: 24070540]
14. Daniels KG, Suo Y, Oas TG. Conformational kinetics reveals affinities of protein conformational states. *PNAS*. 2015; 112:9352–57. [PubMed: 26162682]
15. Daniels KG, Tonthat NK, McClure DR, Chang Y-C, Liu X, et al. Ligand concentration regulates the pathways of coupled protein folding and binding. *J Am Chem Soc*. 2014; 136:822–25. [PubMed: 24364358]
16. Darling RJ, Kuchibhotla U, Glaesner W, Micanovic R, Witcher DR, Beals JM. Glycosylation of erythropoietin affects receptor binding kinetics: role of electrostatic interactions. *Biochemistry*. 2002; 41:14524–31. [PubMed: 12463751]
17. Debye P. Reaction rate in ionic solutions. *J Electrochem Soc*. 1942; 82:265–72.
18. DeVore MS, Braimah A, Benson DR, Johnson CK. Single-molecule FRET states, conformational interchange, and conformational selection by dye labels in calmodulin. *J Phys Chem B*. 2016; 120:4357–64. [PubMed: 27111039]
19. Dogan J, Gianni S, Jemth P. The binding mechanisms of intrinsically disordered proteins. *Phys Chem Chem Phys*. 2014; 16:6323–31. [PubMed: 24317797]
20. EW, Vanden-Eijnden E. Towards a theory of transition paths. *J Stat Phys*. 2006; 123:503–23.
21. Faradjian AK, Elber R. Computing time scales from reaction coordinates by milestoning. *J Chem Phys*. 2004; 120:10880–89. [PubMed: 15268118]
22. Formanek MS, Ma L, Cui Q. Reconciling the “old” and “new” views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). *Proteins*. 2006; 63:846–67. [PubMed: 16475196]
23. Gabdoulline RR, Wade RC. Simulation of the diffusional association of barnase and barstar. *Biophys J*. 1997; 72:1917–29. [PubMed: 9129797]
24. Galburt EA, Rammohan J. A kinetic signature for parallel pathways: conformational selection and induced fit. Links and disconnects between observed relaxation rates and fractional equilibrium flux under pseudo-first-order conditions. *Biochemistry*. 2016; 55:7014–22. [PubMed: 27992996]
25. Getzoff ED, Cabelli DE, Fisher CL, Parge HE, Viezzoli MS, et al. Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature*. 1992; 358:347–51. [PubMed: 1353610]
26. Gianni S, Dogan J, Jemth P. Distinguishing induced fit from conformational selection. *Biophys Chem*. 2014; 189:33–39. [PubMed: 24747333]
27. Gorfe AA, Chang C-E, Ivanov I, McCammon JA. Dynamics of the acetylcholinesterase tetramer. *Biophys J*. 2008; 94:1144–54. [PubMed: 17921202]
28. Gracia B, Xue Y, Bisaria N, Herschlag D, Al-Hashimi HM, Russell R. RNA structural modules control the rate and pathway of RNA folding and assembly. *J Mol Biol*. 2016; 20:3972–85.

29. Greives N, Zhou H-X. BDflex: a method for efficient treatment of molecular flexibility in calculating protein-ligand binding rate constants from Brownian dynamics simulations. *J Chem Phys.* 2012; 137:135105. [PubMed: 23039617]
30. Greives N, Zhou H-X. Both protein dynamics and ligand concentration can shift the binding mechanism between conformational selection and induced fit. *PNAS.* 2014; 111:10197–202. [PubMed: 24982141]
31. Grinstead, CM., Snell, JL. *Introduction to Probability.* Providence, RI: Am. Math. Soc; 2012.
32. Gu S, Silva D-A, Meng L, Yue A, Huang XH. Quantitatively characterizing the ligand binding mechanisms of choline binding protein using Markov state model analysis. *PLOS Comput Biol.* 2014; 10:e1003767. [PubMed: 25101697]
33. Guo J, Zhou H-X. Allosteric activation of SENP1 by SUMO1 β -grasp domain involves a dock-and-coalesce mechanism. *eLife.* 2016; 5:e18249. [PubMed: 27576863]
34. Hammes GG, Chang Y-C, Oas TG. Conformational selection or induced fit: a flux description of reaction mechanism. *PNAS.* 2009; 106:13737–41. [PubMed: 19666553]
35. Held M, Metzner P, Prinz JH, Noe F. Mechanisms of protein-ligand association and its modulation by protein mutations. *Biophys J.* 2011; 100:701–10. [PubMed: 21281585]
36. Huang HC, Briggs JM. The association between a negatively charged ligand and the electronegative binding pocket of its receptor. *Biopolymers.* 2002; 63:247–60. [PubMed: 11807752]
37. Jomain JB, Tallet E, Broutin I, Hoos S, van Agthoven J, et al. Structural and thermodynamic bases for the design of pure prolactin receptor antagonists: X-ray structure of Del1-9-G129R-hPRL. *J Biol Chem.* 2007; 282:33118–31. [PubMed: 17785459]
38. Kiefhaber T, Bachmann A, Jensen KS. Dynamics and mechanisms of coupled protein folding and binding reactions. *Curr Opin Struct Biol.* 2012; 22:21–29. [PubMed: 22129832]
39. Kim E, Lee S, Jeon A, Choi JM, Lee H-S, et al. A single-molecule dissection of ligand binding to a protein with intrinsic dynamics. *Nat Chem Biol.* 2013; 9:313–18. [PubMed: 23502425]
40. Koh CY, Kim JE, Shibata S, Ranade RM, Yu M, et al. Distinct states of methionyl-tRNA synthetase indicate inhibitor binding by conformational selection. *Structure.* 2012; 20:1681–91. [PubMed: 22902861]
41. Koshland DE. Application of a theory of enzyme specificity to protein synthesis. *PNAS.* 1958; 44:98–104. [PubMed: 16590179]
42. Kurisaki I, Takayanagi M, Nagaoka M. Combined mechanism of conformational selection and induced fit in U1A–RNA molecular recognition. *Biochemistry.* 2014; 53:3646–57. [PubMed: 24828852]
43. Lin Y-L, Meng YL, Jiang W, Roux B. Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. *PNAS.* 2013; 110:1664–69. [PubMed: 23319661]
44. Lu H, Tonge PJ. Drug-target residence time: critical information for lead optimization. *Curr Opin Chem Biol.* 2010; 14:467–74. [PubMed: 20663707]
45. Luty BA, El Amrani S, McCammon JA. Simulation of the bimolecular reaction between superoxide and superoxide dismutase: synthesis of the encounter and reaction steps. *J Am Chem Soc.* 1993; 115:11874–77.
46. Luty BA, McCammon JA. Simulation of bimolecular reactions: synthesis of the encounter and reaction steps. *Mol Simul.* 1993; 10:61–65.
47. Luty BA, McCammon JA, Zhou H-X. Diffusive reaction rates from Brownian dynamics simulations: replacing the outer cutoff surface by an analytical treatment. *J Chem Phys.* 1992; 97:5682–86.
48. Ma B, Kumar S, Tsai C-J, Nussinov R. Folding funnels and binding mechanisms. *Protein Eng.* 1999; 12:713–20. [PubMed: 10506280]
49. Majek P, Elber R. Milestoning without a reaction coordinate. *J Chem Theory Comput.* 2010; 6:1805–17. [PubMed: 20596240]
50. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G. String method in collective variables: minimum free energy paths and isocommittor surfaces. *J Chem Phys.* 2006; 125:24106. [PubMed: 16848576]

51. McCammon JA, Northrup SH. Gated binding of ligands to proteins. *Nature*. 1981; 293:316–17. [PubMed: 7278989]
52. Metzner P, Schutte C, Vanden-Eijnden E. Transition path theory for Markov jump processes. *Multiscale Model Simul*. 2009; 7:1192–219.
53. Meyer-Almes FJ. Discrimination between conformational selection and induced fit protein–ligand binding using Integrated Global Fit analysis. *Eur Biophys J*. 2016; 45:245–57. [PubMed: 26538331]
54. Noe F, Schutte C, Vanden-Eijnden E, Reich L, Weikl TR. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *PNAS*. 2009; 106:19011–16. [PubMed: 19887634]
55. Northrup SH, Allison SA, McCammon JA. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J Chem Phys*. 1984; 80:1517–24.
56. Northrup SH, Boles JO, Reynolds JC. Brownian dynamics of cytochrome c and cytochrome c peroxidase association. *Science*. 1988; 241:67–70. [PubMed: 2838904]
57. Pang X, Qin S, Zhou H-X. Rationalizing 5000-fold differences in receptor-binding rate constants of four cytokines. *Biophys J*. 2011; 101:1175–83. [PubMed: 21889455]
58. Pang X, Zhou H-X. Mechanism and rate constants of the Cdc42GTPase binding with intrinsically disordered effectors. *Proteins*. 2016; 84:674–85. [PubMed: 26879470]
59. Pang X, Zhou KH, Qin S, Zhou H-X. Prediction and dissection of widely-varying association rate constants of actin-binding proteins. *PLOS Comput Biol*. 2012; 8:e1002696. [PubMed: 23055910]
60. Paul F, Weikl TR. How to distinguish conformational selection and induced fit based on chemical relaxation rates. *PLOS Comput Biol*. 2016; 12:e1005067. [PubMed: 27636092]
61. Peters GH, Olsen OH, Svendsen A, Wade RC. Theoretical investigation of the dynamics of the active site lid in *Rhizomucor miehei* lipase. *Biophys J*. 1996; 71:119–29. [PubMed: 8804595]
62. Plattner N, Noe F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat Commun*. 2015; 6:7653. [PubMed: 26134632]
63. Qin S, Pang X, Zhou H-X. Automated prediction of protein association rate constants. *Structure*. 2011; 19:1744–51. [PubMed: 22153497]
64. Qin S, Zhou H-X. Dissection of the high rate constant for the binding of a ribotoxin to the ribosome. *PNAS*. 2009; 106:6974–79. [PubMed: 19346475]
65. Rogers JM, Oleinikovas V, Shammas SL, Wong CT, De Sancho D, et al. Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein. *PNAS*. 2014; 111:15420–25. [PubMed: 25313042]
66. Romanowska J, Kokh DB, Fuller JC, Wade RC. Computational approaches for studying drug binding kinetics. In: Keserü, GM, Swinney, DC, editors. *Thermodynamics and Kinetics of Drug Binding*. Vol. 65. Weinheim, Ger: Wiley-VCH Verlag GmbH & Co. KGaA; 2015. p. 211–35.
67. Satpati P, Simonson T. Conformational selection through electrostatics: free energy simulations of GTP and GDP binding to archaeal initiation factor 2. *Proteins*. 2012; 80:1264–82. [PubMed: 22275120]
68. Schneider EV, Böttcher J, Huber R, Maskos K, Neumann L. Structure–kinetic relationship study of CDK8/CycC specific compounds. *PNAS*. 2013; 110:8081–86. [PubMed: 23630251]
69. Schreiber G, Haran G, Zhou H-X. Fundamental aspects of protein–protein association kinetics. *Chem Rev*. 2009; 109:839–60. [PubMed: 19196002]
70. Shan YB, Kim ET, Eastwood MP, Dror RO, Seeliger MA, Shaw DE. How does a drug molecule find its target binding site? *J Am Chem Soc*. 2011; 133:9181–83. [PubMed: 21545110]
71. Sines JJ, Allison SA, McCammon JA. Point-charge distributions and electrostatic steering in enzyme substrate encounter: Brownian dynamics of modified copper-zinc superoxide dismutases. *Biochemistry*. 1990; 29:9403–12. [PubMed: 2248953]
72. Smoluchowski MV. Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Z Phys Chem*. 1917; 92:129–68.
73. Steiner S, Caflisch A. Peptide binding to the PDZ3 domain by conformational selection. *Proteins*. 2012; 80:2562–72. [PubMed: 22753014]

74. Suddala KC, Wang J, Hou Q, Water NG. Mg^{2+} shifts ligand-mediated folding of a riboswitch from induced-fit to conformational selection. *J Am Chem Soc.* 2015; 137:14075–83. [PubMed: 26471732]
75. Sullivan SM, Holyoak T. Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *PNAS.* 2008; 105:13829–34. [PubMed: 18772387]
76. Swift RV, McCammon JA. Substrate induced population shifts and stochastic gating in the PBCV-1 mRNA capping enzyme. *J Am Chem Soc.* 2009; 131:5126–33. [PubMed: 19301911]
77. Szabo A. Theory of diffusion-influenced fluorescence quenching. *J Phys Chem.* 1989; 93:6929–39.
78. Szabo A, Shoup D, Northrup SH, McCammon JA. Stochastically gated diffusion-influenced reactions. *J Chem Phys.* 1982; 77:4484–93.
79. Tan RC, Truong TN, McCammon JA, Sussman JL. Acetylcholinesterase: electrostatic steering increases the rate of ligand binding. *Biochemistry.* 1993; 32:401–3. [PubMed: 8422348]
80. Tsai C-J, Liu S, Hung C-L, Jhong S-R, Sung T-C, Chiang Y-W. BAX-induced apoptosis can be initiated through a conformational selection mechanism. *Structure.* 2015; 23:139–48. [PubMed: 25497728]
81. Vanden-Eijnden E, Venturoli M. Markovian milestoning with Voronoi tessellations. *J Chem Phys.* 2009; 130:194101. [PubMed: 19466815]
82. Vogt AD, Di Cera E. Conformational selection is a dominant mechanism of ligand binding. *Biochemistry.* 2013; 52:5723–29. [PubMed: 23947609]
83. Votapka LW, Amaro RE. Multiscale estimation of binding kinetics using Brownian dynamics, molecular dynamics and milestoning. *PLOS Comput Biol.* 2015; 11:e1004381. [PubMed: 26505480]
84. Wade RC, Davis ME, Luty BA, Madura JD, McCammon JA. Gating of the active site of triose phosphate isomerase: Brownian dynamics simulations of flexible peptide loops in the enzyme. *Biophys J.* 1993; 64:9–15. [PubMed: 8431552]
85. Wade RC, Gabdoulline RR, Ludemann SK, Lounnas V. Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulations. *PNAS.* 1998; 95:5942–49. [PubMed: 9600896]
86. Wade RC, Luty BA, Demchuk E, Madura JD, Davis ME, et al. Simulation of enzyme-substrate encounter with gated active sites. *Nat Struct Biol.* 1994; 1:65–69. [PubMed: 7656010]
87. Wang W, Ye W, Yu Q, Jiang C, Zhang J, et al. Conformational selection and induced fit in specific antibody and antigen recognition: SPE7 as a case study. *J Phys Chem B.* 2013; 117:4912–23. [PubMed: 23548180]
88. Wang Y, Shen B-J, Sebald W. A mixed-charge pair in human interleukin 4 dominates high-affinity interaction with the receptor alpha chain. *PNAS.* 1997; 94:1657–62. [PubMed: 9050834]
89. Whittier SK, Hengge AC, Loria JP. Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases. *Science.* 2013; 341:899–903. [PubMed: 23970698]
90. Wlodek ST, Antosiewicz J, Briggs JM. On the mechanism of acetylcholinesterase action: the electrostatically induced acceleration of the catalytic acylation step. *J Am Chem Soc.* 1997; 119:8159–65.
91. Wu H, Paul F, Wehmeyer C, Noe F. Multiensemble Markov models of molecular thermodynamics and kinetics. *PNAS.* 2016; 113:E3221–30. [PubMed: 27226302]
92. Ye L, Van Eps N, Zimmer M, Ernst OP, Prosser RS. Activation of the A2A adenosine G-protein-coupled receptor by conformational selection. *Nature.* 2016; 533:265–68. [PubMed: 27144352]
93. Yu T-Q, Lapelosa M, Vanden-Eijnden E, Abrams CF. Full kinetics of CO entry, internal diffusion, and exit in myoglobin from transition-path theory simulations. *J Am Chem Soc.* 2015; 137:3041–50. [PubMed: 25664858]
94. Yu X, Martinez M, Gable AL, Fuller JC, Bruce NJ, et al. webSDA: a web server to simulate macromolecular diffusional association. *Nucleic Acids Res.* 2015; 43:W220–24. [PubMed: 25883142]
95. Zhou H-X. Kinetics of diffusion-influenced reactions studied by Brownian dynamics. *J Phys Chem.* 1990; 94:8794–800.

96. Zhou H-X. On the calculation of diffusive reaction rates using Brownian dynamics simulations. *J Chem Phys.* 1990; 92:3092–95.
97. Zhou H-X. Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein-protein association kinetics. *Biophys J.* 1993; 64:1711–26. [PubMed: 8396447]
98. Zhou H-X. Effect of interaction potentials in diffusion-influenced reactions with small reactive regions. *J Chem Phys.* 1996; 105:7235–37.
99. Zhou H-X. Enhancement of protein-protein association rate by interaction potential: accuracy of prediction based on local Boltzmann factor. *Biophys J.* 1997; 73:2441–45. [PubMed: 9370437]
100. Zhou H-X. From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys J.* 2010; 98:L15–17. [PubMed: 20303846]
101. Zhou H-X. Rate theories for biologists. *Q Rev Biophys.* 2010; 43:219–93. [PubMed: 20691138]
102. Zhou H-X. Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. *PNAS.* 2011; 108:8651–56. [PubMed: 21543711]
103. Zhou H-X. Intrinsic disorder: signaling via highly specific but short-lived association. *Trends Biochem Sci.* 2012; 37:43–48. [PubMed: 22154231]
104. Zhou H-X, Briggs JM, McCammon JA. A 240-fold electrostatic rate enhancement for acetylcholinesterase-substrate binding can be predicted by the potential within the active site. *J Am Chem Soc.* 1996; 118:13069–70.
105. Zhou H-X, Pang X, Lu C. Rate constants and mechanisms of intrinsically disordered proteins binding to structured targets. *Phys Chem Chem Phys.* 2012; 14:10466–76. [PubMed: 22744607]
106. Zhou H-X, Szabo A. Theory and simulation of stochastically-gated diffusion-influenced reactions. *J Phys Chem.* 1996; 100:2597–604.
107. Zhou H-X, Szabo A. Theory and simulation of the time-dependent rate coefficients of diffusion-influenced reactions. *Biophys J.* 1996; 71:2440–57. [PubMed: 8913584]
108. Zhou H-X, Szabo A. Enhancement of association rates by nonspecific binding to DNA and cell membranes. *Phys Rev Lett.* 2004; 93:178101. [PubMed: 15525128]
109. Zhou H-X, Wlodek ST, McCammon JA. Conformation gating as a mechanism for enzyme specificity. *PNAS.* 1998; 95:9280–83. [PubMed: 9689071]
110. Zhou H-X, Wong K-Y, Vijayakumar M. Design of fast enzymes by optimizing interaction potential in active site. *PNAS.* 1997; 94:12372–77. [PubMed: 9356456]
111. Zuckerman DM, Chong LT. Weighted ensemble simulation: review of methodology, applications, and software. *Annu Rev Biophys.* 2017; 46:43–57. [PubMed: 28301772]

Figure 1.

Models of protein–ligand binding. (a) A spherical protein surrounded by point-like ligands. The binding reaction occurs as soon as a ligand comes into contact with the protein; this scenario is modeled mathematically by an absorbing boundary condition on the protein surface. (b) The protein undergoing indifferent switch between two conformations. The interconversion rates between the two conformations are the same regardless of where the ligands are. In the inactive conformation (*gray sphere*), the binding reaction cannot proceed; this situation is modeled by a reflecting boundary condition on the protein surface. The active conformation (*cyan sphere*) allows the binding reaction to proceed, as modeled by the absorbing boundary condition. (c) Alternative pathways via which the inactive unbound protein reaches the native complex. In the conformational-selection pathway (*blue arrows*), the protein first makes the inactive-to-active transition and then loosely binds a ligand. In the induced-fit pathway, the order of these two steps is reversed (*red arrows*). The loosely bound complex converts to the native complex upon further tightening of the protein–ligand interactions (*black arrow*). Note that the energy landscapes of the protein are such that the inactive-to-active transition is uphill prior to loose binding but downhill after loose binding. (d) The protein undergoing induced switch. The interconversion rates depend on whether there are loosely bound ligands. Consistent with the energy landscapes in panel c, the rates favor the inactive conformation when the protein is free of any loosely bound ligands (*left arrows*) but favor the active conformation instead when a ligand is loosely bound (*right arrows*).

Figure 2.

The apparent rate ($\mathcal{K}_{\text{app}}$) of the binding kinetics and the protein–ligand binding rate constant (k_b) for the many-body model of Figure 1*d*. (a) Results for $\mathcal{K}_{\text{app}}$ from many-body simulations (*symbols*) and predicted by the four-state model (*curves*). Adapted from Greives & Zhou (30). (b) Results for k_b predicted by the four-state model and from solving the single protein–ligand pair problem (8, 100).

Figure 3.

The induced-fit fraction Φ_{IF} as a quantitative measure of binding mechanism. (a) Results for Φ_{IF} from many-body simulations (*symbols*) and predicted by the four-state model (*curves*). (b) Dependences of Φ_{IF} on ligand concentration and on conformational interconversion rates shown as a two-dimensional surface. Adapted from Greives & Zhou (30).

Figure 4.

Illustration of four algorithms for calculating protein–ligand binding rate constants. (a) The 1984 algorithm of the McCammon group (55). The ligand, started in the *b* surface, can either be absorbed on the boundary of the binding site or be absorbed on the *q* surface. (b) The 1990 algorithm of Zhou (95). The ligand starts within the binding site and can react with rate γ to form the native complex whenever it is inside the binding site. (c) The algorithm of Luty et al. (45). Absorbing state 1 is the binding site, whereas absorbing state 4 is infinity. Transition probabilities in green and red are obtained from simulations where the protein is treated as rigid and flexible, respectively. (d) The BDflex algorithm of Greives & Zhou (29). The dividing surface S^* is absorbing for the exterior problem and partially absorbing for the interior problem. The protein is treated as rigid and flexible, respectively, in the exterior and interior problems.